Executive Summary

I have performed a 3-month stay at NIOO-KNAW. The aim of the stay was to study the horizontal transmission of an important crop pathogen, tobacco mild green mosaic virus (TMGMV), and the effect of satellite tobacco mosaic virus (STMV) on transmission. STMV has a very small genome and is dependent on TMGMV-infected cells for replication. Statistical analyses of experimental data showed a sudden stop in TMGMV transmission in the presence of STMV in a row of tobacco plants. Mathematical modelling of infection spread was performed to describe these trends quantitatively and explore the possible underlying mechanisms. First, we fitted simple mathematical models to the transmission data to describe the patterns of spread. Model results showed that two TMGMV genotypes with distinct biological properties do not differ significantly in their transmission, while the presence of the satellite resulted in a significant decrease in transmission over time. Second, we expanded a standard epidemiological model – the Susceptible Exposed Infectious Removed (SEIR) model – to describe our setup with two virus variants and understand why STMV affects transmission. The fitted model strongly suggests that STMV transmission initially lags behind that of TMGMV, but eventually STMV catches up and strongly reduces TMGMV transmission. These results help explain the patterns I have observed and formulate mechanistic, testable hypotheses for the spread of TMGMV and STMV that I will test in my final year of PhD work.

During the three-month stay, I learned a lot about how to analyze experimental data with mathematical models, including developing models, fitting them to data and coding. I actively participated in activities organized by the Department of Microbial Ecology, including weekly seminars. This stay also allowed for the opportunity to contribute to the Second Annual Workshop of the Netherlands Association for Virus Ecology (October 5th, 2023) and give a seminar in the Virology department of Wageningen University & Research (October 9th, 2023). In both instances this work was discussed, and ISME funding and contribution was acknowledged.

Introduction

Viruses are the main group of emerging pathogens in plants, but the lack of knowledge of virus epidemiology in wild reservoirs limits our understanding of this phenomenon in crops. One of the few cases in which the wild reservoir of a virus is known, is the system formed by tobacco mild green mosaic virus (TMGMV, tobamovirus) and *Nicotiana glauca* Grah. *N. glauca* has been shown to be the reservoir host from which TMGMV spills over into tomato plants cultivated in greenhouses. Wild populations of this host in Southeastern Spain maintain two different virus variants (Long-3'UTR and Short-3'UTR) without causing overt disease, as well as the satellite tobacco mosaic virus (STMV). STMV is a small, spherical ssRNA virus that depends on his helper virus, TMGMV, for replication.

During the 3-month activity developed at the Netherlands Institute of Ecology (NIOO-KNAW), we aimed to develop mathematical models to study the dynamics of infection of both TMGMV genotypes, their maintenance in wild populations of their wild reservoir and the influence of STMV on the process. Prior to the activity, I performed experiments to obtain the data required to develop these models, inspired by previous work from the Department of Microbial Ecology (Ampt et al. 2023. *New Phytologist* **233**: 1303-1316). These experiments consisted of a horizontal transmission study that was performed with one Long-3' UTR (96/5), one Short-3' UTR (99/16) and one Long-3' UTR + STMV (94/5) isolates of TMGMV. In previous work, I have shown that the long and short isolates have different biological properties, such as patterns of viral accumulation (the number of virus genomes present in the sample) over time. For each treatment, ten *Nicotiana tabacum* Samsun plants were placed in horizontal containers, separated by 5 cm, and only the first plant of each row of plants was inoculated with one of the three virus isolates. Four replicates were performed for each virus isolate, and infection

progress was monitored by symptom apparition in the row of plants, and using additional test plants in which a strong plant immune response indicates virus presence. These test plants are a resistant host of TMGMV (*N. tabacum* Xanthi n.c.) which develop of necrotic local lesions when inoculated with tissue containing TMGMV. Quantification of viral accumulation was performed at the inoculated plants at 7 days pot inoculation (dpi) from inoculated leaves and at 7 and 12 dpi for upper leaves to confirm the presence of TMGMV and STMV.

Statistical analysis of infection progress data

We started by performing statistical analyses to look for the main trends of interest in the data, observing that the infection progressed differently through the different rows of plants, showing a sudden stop of the transmission of TMGMV after the third plant in the case of the Long-3'UTR + STMV treatment.

Viral accumulation results show that the treatment with STMV has a significantly lower viral accumulation than the Long-3'UTR treatment in inoculated leaves (t_7 = 3.030, P = 0.008), but this does not occur in upper leaves at 7 dpi (t_8 = 5.933, P = 0.654) nor at 12 dpi (t_8 = 5.987, P = 0.073). Regarding the Short-3'UTR treatment, viral accumulation was significantly lower than Long-3'UTR in the case of the upper leaves (7 dpi: t_8 = 7.250, P = 0.001, 12 dpi: t_8 = 12.874, P = 0.000), but not in inoculated leaves (t_8 = 2.447, P = 0.092). AUDPC values were calculated from symptoms and necrotic local lesion data to evaluate each row of plants as a time-course. In the case of symptom apparition, no significant differences were observed between any of the three treatments (W = 1.365, P = 0.304), but in the case of the necrotic local lesions significant differences were observed between Long-3'UTR and Long-3'UTR + STMV (t_8 = 2.529, P = 0.046) in the progress of infection.

Analyzing the data with a simple mathematical model

Once we had an overview of the data and the trends of each treatment, we aimed to propose a mathematical model that would capture the observed horizontal transmission of TMGMV in each treatment.

First, we fitted a geometric model to the data of the furthest infected plant in each row (y) over time in days post inoculation of the first plant (t). Linear models were immediately discarded, given the strong decrease in virus spread over time, and different variants of a geometric model were evaluated. The geometric model used was:

$$y_{t+1} = y_t + ab^{t-c}.$$

where a is the initial rate of transmission, b determines the rate at which transmission attenuates over time and c is the latent period of the first plant of the line. We can then determine the furthest plant at t, such that when $b \neq 1$ then $y_t = y_0 + a(1 + b^{t-c})/(1 + b)$, and when $b = 1$ then $y_t = y_0 + at$. The starting value (i.e., y_0) is always one, as the first plant is always infected in every row of plants. Values for free model parameters a, b and c were estimated using stochastic hill climbing (SHC), an evolutionary algorithm for model fitting implemented in a custom R script, to minimize the sum of squares. One-hundred repetitions of the search, each starting in a random point in a broad parameter space, were used to ensure a global solution was found. The model was fitted to 1000 bootstrapped datasets to determine the 95% confidence intervals of the parameter estimates, as shown in **Table 1**. There are significant differences between the parameter estimates of the different experiments in the case of the comparison with Long-3' UTR + STMV (parameters a and b), but not in between Long-3' UTR and Short-3' UTR. These model parameter estimates therefore confirm that the rate of transmission is initially higher when STMV is present (as indicated by a significantly higher value of

parameter *a*), while also attenuating more rapidly (as indicated by a significantly lower value of parameter *b*).

As confirmation of these results, we also fitted a stochastic variant of the same model (Model 2 in Ampt et al. 2023. *New Phytologist* **233**: 1303-1316). This model allows stochastic variation in whether a new plant is infected each day, but is the same underlying model. This model therefore predicts not only the mean of y_t, but also its distribution over plants in the row. Rather than using the sum of squares to evaluate the goodness of fit, we can then use the negative log likelihood (NLL) calculated from the multinomial likelihoods obtained by comparing models and data for each time point. This change leads to a more robust model fitting, because the model prediction is more comprehensive and model fitting does not assume normally distributed errors. Fitting the model does require considerable computational resources, and we therefore could not estimate confidence intervals with bootstrapping. The results are shown in **Table 1**, and coincide with the previous strategy, confirming the correct estimation of the model parameters and a minimal effect of model assumptions.

Table 1: Parameter estimates from different modelling strategies are reported: the simple geometric model ("Deterministic model") and the stochastic geometric model ("Stochastic model"). For the deterministic model, 95% confidence intervals after 1000 bootstraps are given in brackets.

Experiment	**Model parameter**	**Model Parameter estimate**	
		Deterministic model	Stochastic model
Long-3'UTR	a	0.59 {0.52, 0.77}	0.71
	b	0.85 {0.62, 0.90}	0.84
	c	4.05 {3.76, 4.35}	4.18
Short-3'UTR	a	0.58 {0.45, 0.83}	0.59
	b	0.88 {0.72, 0.93}	0.89
	c	3.53 {2.64, 4.12}	3.01
Long-3'UTR + STMV	a	1.23 {1.06, 1.48}	1.45
	b	0.30 {0.05, 0.41}	0.54
	c	3.54 {3.30, 3.86}	3.76

Analyzing the data with an epidemiological model

The results obtained until this point support the conclusion that the treatment with STMV behaves differently than the treatments without it, having a significantly higher rate of transmission (*a*) and a significantly higher rate of transmission attenuation (*b*). Furthermore, viral accumulation differences between the treatments do not seem to affect the progress of the horizontal transmission, since Long-3'UTR and Short-3'UTR behave similarly, and it is only the treatment with STMV that shows a sudden stop of the transmission of TMGMV. Thus, an epidemiology model (**Figure 1**) was developed with the goal of exploring what mechanisms might underlie the sudden drop in transmission over time when the satellite is present.

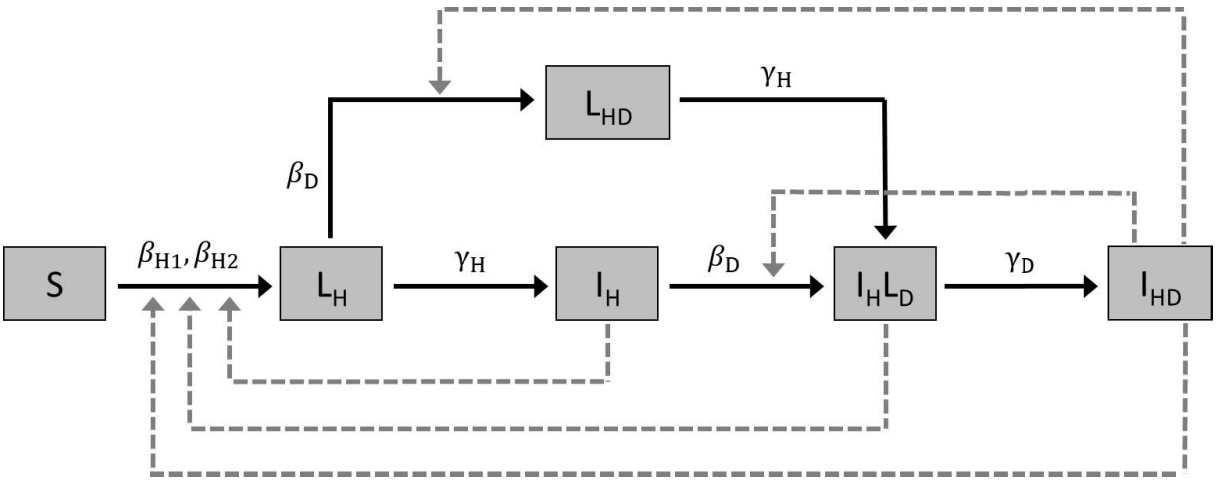

Figure 1: Overview of the model of virus transmission proposed. Boxes with letters represent host states dark lines represent transitions between host states, and grey dotted lines represent an effect of a host state on a transition. Subscripts indicate whether the states (**S**usceptible, **L**atent and **I**nfectious) or model parameters (β and γ) correspond to the helper (H), satellite (D) or both.

In this model, there is a helper virus, TMGMV (H), that is capable of autonomous transmission. In our experiments, this would correspond with, for example, the Long 3' UTR variant of TMGMV. There is also a satellite STMV (D) that can only replicate and be transmitted in the presence of the helper. Based on the presence of these two variants in the system, the following host states are possible: S (susceptible hosts, no H or D present), L_H (H in latent state, a plant infected by H but not yet capable of transmitting this variant), I_H (infectious for H only), I_HL_D (infectious for H, in latent state for D), I_{HD} (infectious for both H and D), and finally the state L_{HD} (a host in L_H state is exposed to D before it transitions to I_H). The fact that these hosts pass through an exposed period twice represents a cost of the satellite on the replication of the helper. Note that the L state is often referred to as the "exposed" state in epidemiology, but we prefer to use the term "latent" to indicate clearly that host has become irreversibly infected. The parameter β is the transmission rate, whereas $1/\gamma$ is the mean time in a latent state that transitions spontaneously to the corresponding infectious state. Subscripts denote which transitions the states and constants apply to, with β_{H1} corresponding to the effects of the I_H and I_HL_D on the transition from L_H to I_{HD}, and β_{H2} corresponding to the effects of the I_{HD} on this transition.

This model was implemented both with and without the consideration that transmission rate decreases over time, reflecting the effect of plant development and/or the dynamic of infection of TMGMV and STMV (parameter k). Implementation of code to generate model predictions has been finalized, and we have also written a first version of a SHC-based routine to estimate model parameters.

We used an epidemiological approach focused on between-host transmission, as STMV did not affect infection characteristics in single plants; the differences caused by the satellite are only manifest when transmission through multiple plants is considered. The comparison with Long and Short 3' UTR variants is striking because here we did not see differences in transmission whereas we did see differences in their biological properties in individual plants. Overall, are approach highlights the importance of scaling up to the between-host level when trying to understand the ecology of viruses.

ISME funds

The funds that ISME approved to grant for the development of this 3-month project at NIOO-KNAW included:

- Bench fee: covered by the host group at NIOO-KNAW.
- Travel: NS flex subscription €1061.40 (3 months x €353.80 per month).

As a part of the project, a contribution to the Second Annual Workshop of the Netherlands Association for Virs Ecology (October 5th, 2023) was given in the form of a talk in which this work was described, and ISME funding and contribution was acknowledged. Moreover, a seminar was given on the matter at the Virology department of Wageningen University & Research (October 9th, 2023). **Figure 2** shows the acknowledgement to ISME in those talks.

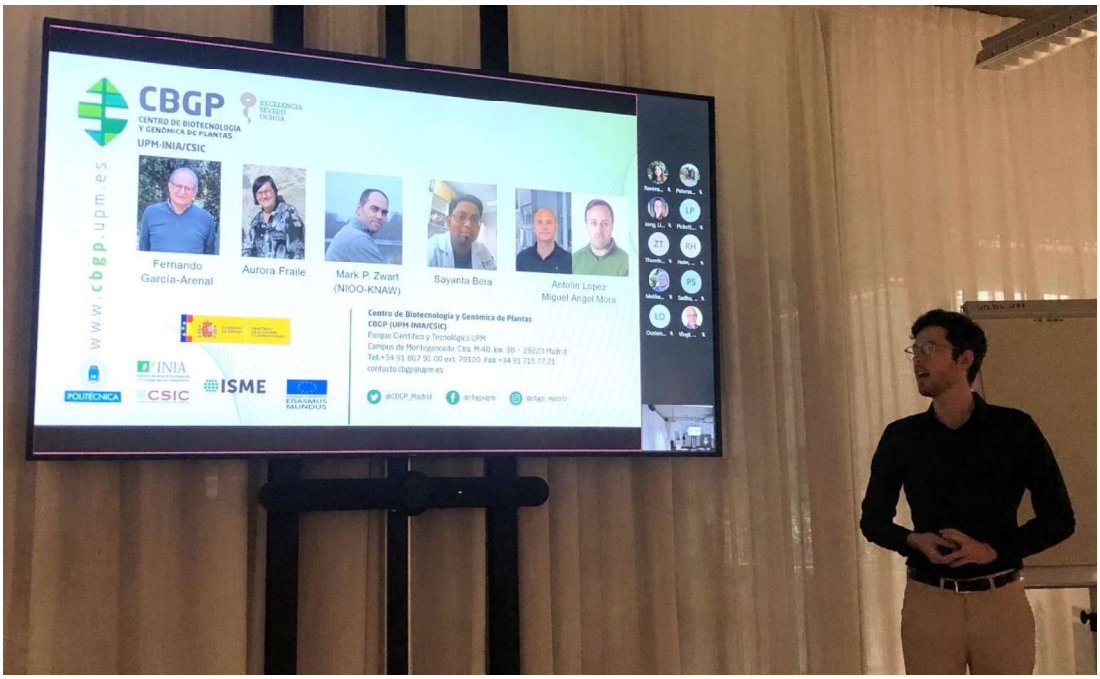

Figure 2: Presentation of the work performed and acknowledgement to ISME. In this presentation, background, motivations and results were discussed. It was given at the Second Annual Workshop of the Netherlands Association for Virus Ecology and at Wageningen University & Research.